Properties of knowledge base and firm survival: Evidence from a sample of French manufacturing firms

Alessandra Colombelli, Jackie Krafft, Francesco Quatraro

Properties of knowledge base and firm survival: Evidence from a sample of French manufacturing firms[1]

Alessandra Colombelli*[a,b] – Jackie Krafft[b] – Francesco Quatraro[b*]

a) Dipartimento di Ingegneria Gestionale e della Produzione, Politecnico di Torino
b) University of Nice Sophia Antipolis and CNRS-GREDEG

ABSTRACT. The paper analyzes the effects of the properties of firms' knowledge base on the survival likelihood of firms. Drawing upon the analysis of the patterns of co-occurrence of technological classes in patent applications, we derive the coherence, variety and cognitive distance indexes, accounting respectively for technological complementarity, differentiation and dissimilarity in the firms' patent portfolios. The results of our analysis are in line with the previous literature, showing that innovation enhances the survival likelihood of firms. In addition, we show that the search strategies at work in the development of firms' knowledge base matter in reducing the likelihood of a failure event. Knowledge coherence and variety appear to be positively related to firms' survival, while cognitive distance exerts a negative effect. We conclude that firms able to exploit the accumulated technological competences have more chances to be successful in competing durably in the market arena, and derive some policy implications concerning the role of public intervention in the orientation of search efforts in local contexts.

JEL Classification codes: O32, L10, L20

Keywords: Knowledge coherence, variety, cognitive distance, firms' survival

*Corresponding Author: Alessandra Colombelli, Dipartimento di Ingegneria Gestionale e della Produzione, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy. Tel.: +39 011 0907275. Fax: +39 011 0907299. Email: alessandra.colombelli@polito.it.

[1] A preliminary version of the paper has been presented to the conference "The governance of a complex world" held in Nice, 1-3 November 2012. The authors wish to thank Fabio Montobbio and Bulat Sanditov for their comments. The authors acknowledge the financial support of the European Union D.G. Research with the Grant number 266959 to the research project 'Policy Incentives for the Creation of Knowledge: Methods and Evidence' (PICK-ME), within the context Cooperation Program / Theme 8 / Socio-economic Sciences and Humanities (SSH).

1 Introduction

The mechanisms underlying the process of new firms' creation and their post-entry performances have been the focus of a wide body of theoretical and empirical literature. Indeed, the intrinsic heterogeneity of firms, as well as of the sectors and the regions in which they operate, generates quite differentiated post-entry dynamics (Santarelli and Vivarelli, 2007).

Most of the analyses carried out in industrial economics provides explanations of the survival patterns of firms based on firms' age and size. Some others also stress the influence of the economic environment, and hence of the geographical localization of firms.

The role played by innovation has been addressed by Audretsch (1991) and Audretsch and Mahmood (1994), by showing that firms' survival rates change according to the belonging to innovative or non-innovative industries. More recently, a number of contributions based on the product lifecycle approach have investigated the effects of innovation on the patterns of firms survival (Cefis and Marsili, 2005 and 2006). The data used in these works are drawn from the Community Innovation Surveys (CIS) and concern (self-reported) innovation efforts of firms, distinguishing between the introduction of product and process innovations. While these studies provide important evidence, Helmers and Rogers (2010) raise some concerns on their utilization due to the possible biases generated by self-reported innovation measures. Instead they use measures related to intellectual property assets of the firms, like patents and trademarks, confirming that innovation enhances the likelihood of firms' survival.

However, the existing literature on innovation and firm survival tends to treat new technologies as a sort of black boxes. When one tries to go beyond self-reported measures, the proxies used in the analysis boil down to some dummies related to the observation of a patent or a trademark in the intangible assets of the firm.

In this paper we aim at providing an empirical account of the role played by the knowledge creation process in the dynamics of firm survival. In particular we analyze the effects of different kinds of firms' search behaviors in the technological landscape by drawing upon a collective knowledge approach. In this perspective new knowledge is the outcome of a process in which different knowledge inputs dispersed in the economy are combined together. The degree of complementarity and similarity of the combined bits characterizes the structure of firms' knowledge base and provides useful information on their ability to move across the technology landscape (Krafft, Quatraro and Saviotti, 2009; Quatraro, 2012).

The results of our empirical study on a sample of French manufacturing firms confirm that innovation increases the survival likelihood. Moreover, firms that have developed specific competences to move in areas of the technological landscape which are closer to their accumulated technological competences are more likely to survive than firms that search in areas far from their core competences and hence are subject to high degrees of technological uncertainty.

The rest of the paper is organized as follows. Section 2 articulates the theoretical framework linking the survival patterns of firms with the structural properties of their knowledge bases. Section 3 presents the data, the variables and the methodology. We show the empirical results of the analysis in Section 4 and provide the conclusions and policy implications in Section 5.

2 Firm survival and knowledge creation

A wide body of empirical literature in industrial organization shows that firms' population is characterized by endless turbulence. The dynamics of firms' demography is indeed marked by high rates of turnover, both across and within industries (Caves, 1998). The analysis of entry and exit dynamics have been originally closely intertwined. New firms are indeed exposed to high rates of mortality in the critical start-up period, and hence the characteristics of firms at the moment of their creation have been usually regarded as good predictors of post-entry performances, i.e. on the probability for firms to survive to market selection (Dunne et al., 1988; Audrestch, 1995; Baldwin, 1995).

Out of the factors influencing the failure likelihood, the existing literature identifies two key elements, i.e. size and age. On the one hand, the former is basically related to Gibrat's law of proportionate effects. In this perspective, new firms entering the market have more chances to attenuate post-entry mortality if they are set up on a large scale of production (Sutton, 1997; Mata and Portugal, 1994; Geroski, 1995). On the other hand, the latter is grounded on Jovanovic (1982) theory of 'noisy selection' that explicitly centers the attention on the learning dynamics characterizing firms' behavior. In this framework, firms are not aware of their efficiency level as compared to the general efficiency level of the sector. They discover their efficiency over time, so that those that are relatively efficient survive and grow while those that are inefficient eventually leave the market. The probability of survival hence increases with firms' age. Some non-linear effects have been also observed, according to which the positive relationship is decreasing over time (Evans, 1987; Hall, 1987).

Jovanovic's model has been further extended by Ericson and Pakes (1995), who include firms' investments in R&D so as to make learning the outcome of an intentional choice. By exploring the technological landscape, firms improve their efficiency and profitability and hence their survival likelihood. Rather than the mere effect of time, in this model learning stems from deliberate strategies aimed at enriching their distinctive competences. In the same vein, Nelson and Winter (1982) posit that investments in innovative and imitative R&D on average lead to an improvement of firms productivity levels. The comparison of these latter with the general efficiency level of the sector shapes firms' decision as to whether stay in the market or exit.

On a different and yet complementary ground, Agarwal and Audretsch (2001) show that while size and age are important, their effects on firms' survival change across different sectors according to the stage of the industry lifecycle and the technological regime (Klepper, 1996 and 1997). Size is more likely to matter in the formative stage of an industry, when innovation activities are not yet routinized, than in the mature stages, when innovation activities are rather routinized and small firms can achieve successful strategic positions by filling some market niches that are left empty by incumbents (Caves and Porter, 1977).

Innovation and technological change hence come to the fore in the discussion on firm survival. While the lifecycle approach indirectly addresses the issue by comparing survival patterns across different technological regimes, the theoretical model by Ericson and Pakes and the one by Nelson and Winter establish a direct link between technological efforts and post-entry performances. However, direct empirical assessments of such relationships are still underdeveloped and in any case rely on rather stylized representations of the innovation process. Some investigations (Hall, 1987; Perez et al., 2004) used R&D investments as proxies of innovation activities, by concluding that they are positively related to the survival likelihood of the firm. These are clearly input measures of the innovation process. On the output side, Christensen et al. (1998) analyzed the effects of architectural innovation, while

Banbury and Mitchell (1995) focused their study on the number of product innovations brought about by the firms in their dataset.

More recently, Cefis and Marsili (2005 and 2006) provide an account of the differential effect of the introduction of product and process innovation on firm survival. By linking the results to the product lifecycle theory, according to which the introduction of product innovation characterizes the early stage of the cycle while process innovation becomes more important when the sector comes to maturity, they find that process innovation matters more than product innovation. Helmers and Rogers (2010) adopt an empirical strategy based on intellectual property activity of firms by focusing on a sample of British firms. Intellectual property is proxied in their analysis by patent applications and trademarks, showing that both influence negatively the failure rate of the sampled firms.

While the link between innovation and firm survival seems to be now rather established, little has been said about the importance of search strategies pursued by firms to generate new technological knowledge. The grafting of the recent theories of knowledge creation onto the debate on survival can be far reaching in enhancing the understanding of the differential effects of exploration and exploitation strategies along different stages of the technology lifecycle (March, 1991)[2].

Traditional approaches to technological knowledge have mostly represented it as a homogeneous stock, as if it were the outcome of a quite uniform and fluid process of accumulation made possible by R&D investments, the same way as capital stock (Griliches,

[2] A wide body of literature has recently focused on the issue of ambidexterity, i.e. the optimal balance between exploration and exploitation strategies. In this perspective successful firms are able to achieve an optimal balance between the two dimensions so as to cope both with radical and incremental changes in the technological environment. The appreciation of such issue goes however beyond the scope of this paper.

1979; Mansfield, 1980). Such kind of representation is hardly useful to qualify firms' search behaviors, as it only allows for evaluating it from a quantitative rather than a qualitative viewpoint.

More recently, an increasingly share of scholars in the economics of innovation has elaborated theoretical approaches wherein the process of knowledge production is viewed as the outcome of a recombination process (Weitzmann, 1998; Kauffman, 1993). The creation of new knowledge is represented as a search process across a set of alternative components that can be combined one another. A crucial role is played here by the cognitive mechanisms underlying the search process aimed at exploring the knowledge space so as to identify the pieces that might possibly be combined together. The set of potentially combinable pieces turns out to be a subset of the whole technological space. Search is supposed to be local rather than global, while the degree of localness appears to be the outcome of cognitive, social and technological influences (Saviotti, 2004 and 2007). The ability to engage in a search process within spaces that are distant from the original starting point is likely to generate breakthroughs stemming from the combination of brand new components (Nightingale, 1998; Fleming, 2001).

Based on these achievements, we can introduce the concept of knowledge structure. If knowledge stems from the combination of different technologies, knowledge structure can be represented as a web of connected elements. The nodes of this network stand for the elements of the knowledge space that may be combined with one another, while the links represent their actual combinations. The frequency with which two technologies are combined together provides useful information on how we can characterize the internal structure of the knowledge base. Basically, this characterization takes into account the average degree of complementarity of the technologies which knowledge bases are made of, as well as to the variety of the observed pairs of technologies that lead us to derive three main properties of knowledge structure at a general level:

- **Variety** is related to the technological differentiation within the knowledge base, in particular with respect to the diverse possible combinations of pieces of knowledge in firms' patent portfolios, from the creation of a radically new type of knowledge to the more incremental recombination of already existing types of knowledge.
- **Coherence** can be defined as the extent to which the pieces of knowledge that agents combine to create new knowledge are complementary one another.
- **Similarity** (or dissimilarity) refers to the extent to which the pieces of knowledge used are close one another in the technological space.

The dynamics of technological knowledge can therefore be understood as the patterns of change in its own internal structure, i.e. in the patterns of recombination across the elements in the knowledge space. This approach captures both the cumulative character of knowledge creation, as well as the possible link to the relative stage of development of a technological trajectory (Dosi, 1982; Saviotti, 2004 and 2007; Krafft, Quatraro and Saviotti, 2009).

In this perspective, the generation of new knowledge and the introduction of innovation are the results of cumulative patterns, learning dynamics and path dependence. At the onset of a technological trajectory firms are likely to move in an uncertain environment, so that their innovation efforts are grounded on the recombination of technologies which are characterized by low coherence degree and high cognitive distance. In this phase firms tend to rely more on exploration than exploitation strategies. On the contrary, as the trajectory gets more established, firms have learned both from interactions and from their past activity to identify profitable directions for their search efforts. As time goes by therefore firms tend to privilege the exploitation of the variety of cumulated competences and are more likely to focus their search efforts in well defined areas of the technological landscape. Hence their technology portfolios appear to be featured by high coherence and low degrees of cognitive distance.

The combination of the firm survival framework with the recombinant and collective approach allows us to articulate our working hypotheses concerning the relationship between search strategies and firm survival.

The post-entry performances of firms are characterized by a high degree of turbulence. Besides the environmental factors related to the features of the regions and the sectors in which they operate, some firm-level factors play an important role in shaping the likelihood of survival. In this perspective, technological activities are likely to exert a strong influence on the patterns of exit. Technological knowledge, however, is far from being an undifferentiated bundled stock. On the contrary, it is the result of a combinatorial activity which rests upon the search efforts committed by firms. The collective approach to knowledge creation allow us to propose the concept of knowledge structure, which refers to the shape that features the patterns of recombination.

The structure of knowledge is therefore represented by the elements that are combinable and by the actual observed combinations. Each knowledge bit can be assigned to a technological domain, so that we can characterize the structure of the knowledge base according to the fact that it is made by the combination of more or less similar and complementary elements.

Firms entering in new technological trajectories and exploring the technology landscape, are more likely to undergo failure events due to the high degree of uncertainty that characterizes these turbulent phases and the difficult, length and resource-intensive activity necessary to translate radical breakthroughs into profitable bits of knowledge. On the contrary, firms exploiting their accumulated knowledge, are more likely to be successful, and hence they should show relatively higher survival rates.

3 Data, Variables and Methodology

3.1 The Data

In order to investigate the effects of the properties of knowledge structure on firm survival we gather firm-level data from the Bureau Van Dijk DIANE dataset, which provides detailed information on French firms, and from the PATSTAT database (April 2011), which contains detailed information on worldwide patent applications to the European Patent Office. This information is crucial to implement the properties of knowledge structure that will be described in what follows.

The data obtained from the DIANE dataset refer to a sample of manufacturing firms covering a time span ranging from 2001 (first observed year) to 2011. We decided to focus on manufacturing firms as the use of patents as a proxy for knowledge creation activities clearly raises some concerns when service activities are at stake. We obtained a former dataset of 851,070 firms spread over 36 2-digits NACE industrial sectors (rev2.1). The sectoral distribution of the sample is reported in Table 1, where it is compared to the distribution of firms across the same sectors at the national level.As it is clear from the figures, our sample provide a good approximation of the French sectoral distribution of firms, even though a few sectors seem to be over-represented in our sample.

>>> INSERT Table 1 ABOUT HERE <<<

Since the dataset starts in 2001, we decided to drop from the dataset the firms that have been created after 2001, in order to avoid truncation problems.

We used the harmonized matching tables described by Thoma et al. (2010) to combine the PATSTAT and the DIANE datasets on the basis of the Bureau Van Dijk firm identification code. The harmonization procedure drew on Named Entity Recognition (NER) methods from bioinformatics and applied two different approaches to data integration in the context of patent information. The dictionary-based approach relies on the collection of large datasets of names and their variants, while the rule-based approach articulates a set of rules to establish similarity links across different entity names. The methodology has been then applied by the authors to several data sources, including major patent databases and business directories such as Amadeus[3].

After the merge with the patent datasets and the data cleaning we ended up with 74,862 firms operating in the manufacturing sector. Figure 1 shows the distribution of patent applications across sampled innovating firms. As it is clear, within the subsample of innovative firms, about the 42% of firms hold only one patent, then the 32% hold between 2 and 5 papers, the 10% hold between 6 and 12 papers, while the 12% hold more than 13 papers.

>>> INSERT Figure 1 ABOUT HERE <<<

The final dataset provides firm-level information about economic and innovation activities along the whole observed period. Hence we are able to derive the ‘starting conditions’ for the

[3] More details can be found in the paper by Thoma et al. (2010).

relevant variables at 2002[4]. We are also able to trace the existence of the firm month by month up to December 2010. This is done by using the information on the juridical status of the firm. Following a large body of the literature on the subject, survival data refer here to firms that have failed or have been the object of merger and acquisitions[5] (Agarwal and Audretsch, 2001; Cefis and Marsili, 2006).

3.2 The Variables

3.2.1 Dependent Variable

In order to implement our empirical analysis on survival likelihood we adopted the *survival time* of a firm as key variable. The survival time is calculated for all the firms included in our dataset and extends to all firms of varying ages and sizes. As initial point for the calculation of the survival time we took January 2002. The survival time is therefore the time elapsed between January 2002 and the month in which the firm exited. The survival time is right-censored on December 2011, as an exit event is not observed for continuing firms.

3.2.2 The Key Covariates: Implementation of Knowledge Indicators

The properties of the knowledge base are calculated by using the information contained in patent documents. Since we needed to derive the values of these properties at 2002, we implemented the yearly calculation of the variables described in what follows, and then used the average value on the period 1997-2002.

[4] Due to the need to calculate firm growth as an explanatory control variable.

[5] This can be a severe problem in some contexts, and can be hardly solved with the data we have. However for what concerns the present analysis one may assume that most of M&As involve fast growing innovating firms. This means that we are likely to overestimate failure events for innovative firms. Since our results show that survival rates are positively related to innovation, the assimilation of deaths and M&A leads to an underestimation of the importance of innovation for firms' survival. In other words, were it be possible to distinguish between death and M&A in our data, the results would be even stronger.

For what concerns the definition of the variables, let us start by the traditional firm's *knowledge stock*. This is computed by applying the permanent inventory method to patent applications. We calculated it as the cumulated stock of past patent applications using a rate of obsolescence of 15% per annum: $E_{i,t} = \dot{h}_{i,t} + (1-\delta)E_{i,t-1}$, where $\dot{h}_{i,t}$ is the flow of patent applications and δ is the rate of obsolescence[6].

The implementation of knowledge characteristics proxying for variety, coherence and similarity, rests on the recombinant knowledge approach. In order to provide an operational translation of such variables one needs to identify both a proxy for the bits of knowledge and a proxy for the elements that make their structure. For example one could take scientific publications as a proxy for knowledge, and look either at keywords or at scientific classification (like the JEL code for economists) as a proxy for the constituting elements of the knowledge structure. Alternatively, one may consider patents as a proxy for knowledge, and then look at technological classes to which patents are assigned as the constituting elements of its structure, i.e. the nodes of the network representation of recombinant knowledge. In this paper we will follow this latter avenue[7]. Each technological class *j* is linked to another class *m* when the same patent is assigned to both of them[8]. The higher is the number of patents jointly assigned to classes *j* and *m*, the stronger is this link. Since technological classes attributed to patents are reported in the patent document, we will refer to

[6] A similar approach is used by Soete et Patel (1985).

[7] The limits of patent statistics as indicators of technological activities are well known. The main drawbacks can be summarized in their sector-specificity, the existence of non-patentable innovations and the fact that they are not the only protecting tool. Moreover the propensity to patent tends to vary over time as a function of the cost of patenting, and it is more likely to feature large firms (Pavitt, 1985; Griliches, 1990). Nevertheless, previous studies highlighted the usefulness of patents as measures of production of new knowledge. Such studies show that patents represent very reliable proxies for knowledge and innovation, as compared to analyses drawing upon surveys directly investigating the dynamics of process and product innovation (Acs et al., 2002). Besides the debate about patents as an output rather than an input of innovation activities, empirical analyses showed that patents and R&D are dominated by a contemporaneous relationship, providing further support to the use of patents as a good proxy of technological activities (Hall et al., 1986).

[8] In the calculations 4-digits technological classes have been used.

the link between j and m as the co-occurrence of both of them within the same patent document[9].

On this basis we calculated the following three key characteristics of firms' knowledge bases (see the appendix A for the methodological details):

a) Knowledge variety (KV) measures the degree of technological diversification of the knowledge base. It is based on the informational entropy index.

b) Knowledge coherence (COH) measures the degree of complementarity among technologies.

c) Cognitive distance (CD) expresses the dissimilarities amongst different types of knowledge.

3.2.2.1 Knowledge variety measured by the informational entropy index

Knowledge variety (KV) is measured by using the informational entropy index. Entropy measures the degree of disorder or randomness of the system; systems characterized by high entropy are characterized by high degrees of uncertainty (Saviotti, 1988). Informational entropy has some interesting properties (Frenken and Nuvolari, 2004) including multidimensionality.

Consider a pair of events (X_l, Y_j), and the probability of their co-occurrence p_{lj}. A two dimensional variety measure can be expressed as follows (firm subscripts are omitted throughout the section for the sake of clarity):

[9]It must be stressed that to compensate for intrinsic volatility of patenting behaviour, each patent application is made last five years in order to reduce the noise induced by changes in technological strategy.

$$KV \equiv H(X,Y) = \sum_{l}\sum_{j} p_{lj} \log_2\left(\frac{1}{p_{lj}}\right) \qquad (1)$$

Let the events X_l and Y_j be citation in a patent document of technological classes *l* and *j* respectively. Then p_{lj} is the probability that two technological classes *l* and *j* co-occur within the same patent. The measure of multidimensional entropy, therefore, focuses on the variety of co-occurrences or pairs of technological classes within patent applications[10].

3.2.2.2 The knowledge coherence index

Firms need to combine or integrate many different pieces of knowledge to produce a marketable output. Competitiveness requires new knowledge and knowledge about how to combine old and new pieces of knowledge. We calculate the coherence of firms' knowledge bases, defined as the average relatedness or complementarity of a technology chosen randomly within the firm's patent portfolio with respect to any other technology (Nesta and Saviotti, 2005, 2006; Nesta, 2008)[11].

Obtaining the knowledge coherence index requires a number of steps. First of all, we need to calculate the weighted average relatedness WAR_l of technology *l* with respect to all other technologies in the firm patent portfolio. This measure builds on the measure of *technological*

[10] It must be noted that by measuring the degree of technological differentiation, the calculation of information entropy is affected by the number of technological classes observed, but not necessarily by the number of technological classes in the classification itself. Indeed, the introduction of new technological classes that are not observed does not affect the calculations in that they would be events with zero probability. Entropy rises or falls according to the number of technological classes that are actually observed in the patent sample. It reaches the maximum if all events are equiprobable, i.e. if all technological classes show the same relative frequency. If probabilities are unevenly distributed, one can have very low values of information entropy even if a very large number of technologies is observed.

[11] The function used to measure coherence is completely different from the one used to measure informational entropy. The fact that in both cases the co-occurrence of technological classes enters the calculations does not mean that both functions must lead to the same result. The informational entropy function measures the variety of the set, corresponding to the number of distinguishable entities it contains. The coherence function was introduced by Teece et al (1994) to measure the coherence of a firm based on its products. Nesta and Saviotti (2005, 2006) have subsequently adapted it to measure the coherence of the knowledge base of a firm. The coherence function measures the extent to which the distinguishable entities in the set (in our case the types of knowledge corresponding to different technological classes) are used together irrespective of the number of entities contained in the set. The two functions are in principle independent since they use the same type of data to calculate different properties of the same system. The mathematical independence of the two functions does not imply that the evolution of the corresponding properties is independent. Thus, if new technological classes are introduced into the knowledge base of a sector (an increase in the number of distinguishable entities of the set) there is no reason to expect the capacity of firms to combine the new types of knowledge to be created instantly. We expect that as new types of knowledge are introduced into the knowledge base of a sector, the firms will slowly learn to combine them thus leading to a temporary fall in coherence.

relatedness τ_{lj} (Nesta and Saviotti, 2005, 2006). We start by calculating the relatedness matrix. The technological universe consists of *k* patent applications across all sampled firms. Let P_{lk} = 1 if the patent *k* is assigned the technology *l* [*l*= 1, …, n], and 0 otherwise. The total number of patents assigned to technology *l* is $O_l = \sum_k P_{lk}$. Similarly, the total number of patents assigned to technology *j* is $O_j = \sum_k P_{jk}$. Since two technologies can occur within the same patent, $O_l \cap O_j \neq \varnothing$, and thus the observed the number of observed co-occurrences of technologies *l* and *j* is $J_{lj} = \sum_k P_{lk} P_{jk}$. Applying this relationship to all possible pairs yields a square matrix Ω (n × n) in which the generic cell is the observed number of co-occurrences:

$$\Omega = \begin{bmatrix} J_{11} & & J_{l1} & & J_{n1} \\ \vdots & \ddots & & & \vdots \\ J_{1j} & & J_{lj} & & J_{nj} \\ \vdots & & & \ddots & \vdots \\ J_{1n} & \cdots & J_{\mathrm{ln}} & \cdots & J_{nn} \end{bmatrix} \tag{5}$$

We assume that the number x_{ij} of patents assigned to technologies *i* and *j* is a hypergeometric random variable of the mean and variance:

$$\mu_{lj} = E(X_{lj} = x) = \frac{O_i O_j}{K} \tag{6}$$

$$\sigma_{lj}^2 = \mu_{lj} \left(\frac{K - O_l}{K} \right) \left(\frac{K - O_j}{K - 1} \right) \tag{7}$$

If the observed number of co-occurrences J_{ij} is larger than the expected number of random co-occurrences μ_{ij}, then the two technologies are closely related: the fact that the two technologies occur together in the number of patents x_{ij} is not common or frequent. Hence, the measure of relatedness is given by the difference between the observed and the expected numbers of co-occurrences, weighted by their standard deviation:

$$\tau_{lj} = \frac{J_{lj} - \mu_{lj}}{\sigma_{lj}} \tag{8}$$

Note that this measure of relatedness has no lower or upper bounds: $\tau_{lj} \in \left]-\infty;+\infty\right[$. Moreover, the index shows a distribution similar to a t-test, so that if $\tau_{lj} \in \left]-1.96;+1.96\right[$, we can safely assume the null hypothesis of non-relatedness of the two technologies *i* and *j*. The technological relatedness matrix Ω' can be considered a weighting scheme to evaluate the technological portfolio of firms.

Following Teece et al. (1994), WAR_l is defined as the degree to which technology *l* is related to all other technologies $j \in l$ in the firm's patent portfolio, weighted by patent count P_{jt}:

$$\mathrm{WAR}_{\mathrm{lt}} = \frac{\sum_{\mathrm{j}\neq\mathrm{l}} \tau_{\mathrm{lj}} \mathrm{P}_{\mathrm{jt}}}{\sum_{\mathrm{j}\neq\mathrm{l}} \mathrm{P}_{\mathrm{jt}}} \tag{9}$$

Finally the coherence of the firm's knowledge base at time *t* is defined as the weighted average of the WAR_{lt} measure:

$$COH_{\mathrm{t}} = \sum_{\mathrm{l}} \mathrm{WAR}_{\mathrm{lt}} \times \frac{\mathrm{P}_{\mathrm{lt}}}{\sum_{\mathrm{l}} \mathrm{P}_{\mathrm{lt}}} \tag{10}$$

Note that this index implemented by analysing the co-occurrence of technological classes within patent applications, measures the degree to which the services rendered by the co-occurring technologies are complementary, and is based on how frequently technological classes are combined in use. The relatedness measure τ_{lj} indicates that utilization of technology *l* implies use also of technology *j* in order to perform specific functions that are not reducible to their independent use. This makes the coherence index appropriate for the purposes of this study and marks a difference from entropy, which measures technological differentiation based on the probability distribution of pairs of technological classes across the patent sample[12].

[12] To make it clear, informational entropy is a diversity measure which allows to accounting for variety, i.e. the number of categories into which system elements are apportioned, and balance, i.e. the distribution of system elements across categories. (Stirling, 2007). In this sense entropy does not say anything about the relationships between technological classes, but provides a measure of the diversity of technological co-occurrences, suggesting whether in a sector most of the observed co-occurrences focus on a specific couple or on the contrary whether the observed co-occurrences relate to a large number couples. In this framework, related and unrelated variety provide a measure of the extent to which observed variety applies to couples of technologies that belong to the same macro domain or to different macro-domains. One would expect established technologies to be characterized by relatively low variety of co-occurrences, insofar as the recombination focus on a relatively small numbers of technological classes that have proved to be particularly fertile. On a different ground, the coherence index is based on a normalized measure of how much each observed technology is complementary to

If the coherence index is high, this means that the different pieces of knowledge have been well combined or integrated during the search process. Due to a learning dynamics, firms have increased capability to identify the bits of knowledge that are required jointly to obtain a given outcome. In a dynamic perspective, therefore, increasing values for knowledge coherence are likely to be associated with search behaviours mostly driven by organized search within well identified areas of the technological landscape. Conversely, decreasing values of knowledge coherence are likely to be related to search behaviours mostly driven by random screening across untried areas of the technological landscape in the quest for new and more profitable technological trajectories.

3.2.2.3 The cognitive distance index

We need a measure of cognitive distance (Nooteboom, 2000) to describe the dissimilarities among different types of knowledge. A useful index of distance can be derived from *technological proximity* proposed by Jaffe (1986, 1989), who investigated the proximity of firms' technological portfolios. Breschi et al. (2003) adapted this index to measure the proximity between two technologies[13].

Let us recall that $P_{lk} = 1$ if the patent k is assigned the technology l [l= 1, …, n], and 0 otherwise. The total number of patents assigned to technology l is $O_l = \sum_k P_{lk}$. Similarly, the total number of patents assigned to technology j is $O_j = \sum_k P_{jk}$. We can, thus, indicate the

all other technologies in the analyzed patents. In this sense it cannot be understood as a measure of diversity. The relatedness index indeed provides a measure of the degree to which two technologies are actually jointly used as compared to the expected joint utilization. The index allows to establishing a relationship of complementarity between the technologies in the analyzed patents. Based on the relatedness measure (tau), the coherence index provides an aggregate description of the degree to which the observed technologies in a given sector are complementary to one another.

[13] Cognitive distance is the inverse of similarity or the equivalent of dissimilarity. The measure of similarity has been introduced by biologists and ecologists to measure the similarity of biological species and to understand to what extent they could contribute to biodiversity. The same measure has been applied by Jaffe (1986) to the similarity of technologies. It is not the only possible measure of similarity but it is the most frequently used one. The rational for its use is starts from the assumption that when two technologies, i and j, can be combined with a third technology k, they are similar. We call this measure cognitive distance both because the two terms are used as synonyms in the biological literature and, even more so, because cognitive distance is a concept used by Bart Nooteboom (2000) which has a number of very interesting implications for firm behavior and performance. In particular, the cognitive distance between different firms is expected to affect the probability that they form technological alliances. Intuitively, the need for a firm to learn a completely new technology (discontinuity) will lead to the incorporation into the firm's knowledge base of new patent classes, which would make the knowledge base recognizably different from what it was at previous times. The dissimilarity of the knowledge base can be expected to keep rising with respect to the pre-discontinuity knowledge base until the technology lifecycle has achieved maturity, at which stage the knowledge base of the firm will have stabilized, thus leading to a fall in cognitive distance.

number of patents that are classified in both technological fields *l* and *j* as: $V_{lj} = \sum_k P_{lk} P_{jk}$. By applying this count of joint occurrences to all possible pairs of classification codes, we obtain a square symmetrical matrix of co-occurrences whose generic cell V_{lj} reports the number of patent documents classified in both technological fields *l* and *j*.

Technologiocal proximity is proxied by the cosine index, which is calculated for a pair of technologies *l* and *j* as the angular separation or uncentred correlation of the vectors V_{lm} and V_{jm}. The similarity of technologies l and j can then be defined as follows:

$$S_{lj} = \frac{\sum_{m=1}^{n} V_{lm} V_{jm}}{\sqrt{\sum_{m=1}^{n} V_{lm}^2} \sqrt{\sum_{m=1}^{n} V_{jm}^2}} \tag{11}$$

The idea behind the calculation of this index is that two technologies j and l are similar to the extent that they co-occur with a third technology *m*. Such measure is symmetric with respect to the direction linking technological classes, and it does not depend on the absolute size of technological field. The cosine index provides a **measure of the similarity between two technological fields in terms of their mutual relationships with all the other fields**. S_{lj} is the greater the more two technologies *l* and *j* co-occur with the same technologies. It is equal to one for pairs of technological fields with identical distribution of co-occurrences with all the other technological fields, while it goes to zero if vectors V_{lm} and V_{jm} are orthogonal (Breschi et al., 2003)[14]. Similarity between technological classes is thus calculated on the basis of their relative position in the technology space. **The closer technologies are in the technology space, the higher is S_{lj} and the lower their cognitive distance** (Engelsman and van Raan, 1991; Jaffe, 1986; Breschi et al., 2003).

The cognitive distance between j and l can be therefore measured as the complement of their index of technological proximity:

$$d_{lj} = 1 - S_{lj} \tag{12}$$

Having calculated the index for all possible pairs, it needs to be aggregated at the firm level to obtain a synthetic index of distance amongst the technologies in the firm's patent portfolio. This is done in two steps. First we compute the weighted average distance of technology *l*, i.e. the average distance of *l* from all other technologies.

[14] For Engelsman and van Raan (1991), this approach produces meaningful results particularly at a 'macro' level, i.e. for mapping the entire domain of technology.

$$\mathrm{WAD}_{lt} = \frac{\sum_{j \neq l} d_{lj} P_{jt}}{\sum_{j \neq l} P_{jt}} \tag{13}$$

where P_j is the number of patents in which the technology j is observed. The average cognitive distance at time t is obtained as follows:

$$\mathrm{CD}_t = \sum_l \mathrm{WAD}_{lt} \times \frac{P_{lt}}{\sum_l P_{lt}} \tag{14}$$

The cognitive distance index measures the inverse of the similarity degree among technologies. When cognitive distance is high, this is an indication of the increased difficulty or cost the firm faces to learn the new type of knowledge which is located in a remote area of the technological space. Increased cognitive distance is related to the emergence of discontinuities associated with paradigmatic shifts in the sector knowledge base. It signals the combination of core technologies with unfamiliar technologies.

The adoption of these variables marks an important step forward in the operational translation of knowledge creation processes. In particular, they allow for a better appreciation of the collective dimension of knowledge dynamics. Knowledge is indeed viewed as the outcome of a combinatorial activity in which intentional and unintentional exchange among innovating agents provides the access to external knowledge inputs (Fleming and et al., 2007). The network dynamics of innovating agents provide the basis for the emergence of new technological knowledge, which is in turn represented as an organic structure, characterized by elementary units and by the connections amongst them.

An increase in knowledge coherence is likely to signal that firms innovation activities are dominated by the exploitation of accumulated technological competences, while a decrease in knowledge coherence is linked tothe emergence of new technological trajectories which firms address by means of exploration across untried areas of the technology landscape. Similarly, increasing values of cognitive distance are related to   exploration dynamics, while decreasing cognitive distance is more likely to be linked to exploitation and organized search across the technology landscape . Knowledge variety is likely to increase in any case when new combinations are introduced in the system. (Krafft, Quatraro, Saviotti, 2009).

Consistently with the theoretical framework lied down in the previous section, we expect the survival rates of the firms to be positively related to knowledge coherence and knowledge variety, and negatively related to cognitive distance.

3.2.3 Control variables

Besides the effects of the knowledge related variables, we also control for the effects of a number of variables that have proved to affect the survival likelihood in previous empirical settings.

To this purpose we include in the regressors vector the current *size of the firm* at the beginning of the period of observation. The variable is derived by the DIANE dataset and measured as the log of firms' sales at 2002. Moreover, in order to account for possible non-linear effects of size on survival, we also included the squared term of firm size in the econometric estimation (Evans, 1987; Hall, 1987).

Also *firm's age* has been found to affect survival patterns. The age of the firm is calculated in terms of elapsed years since the foundation of the firm. Also in this case, since the relationship between age and survival can be non-linear and take a U-inverted shape, we included the squared term of age in the econometric model (Evans, 1987; Bruderl and Schussler, 1990). Moreover, we also accounted for the possibility for size and age to interact (Cefis and Marsili, 2006).

In addition to size and age, we also accounted for the effects of differential *growth* rates at the beginning of the observed period (Agarwal, 1997). We calculated firm's growth as the log difference of firms sales between 2002 and 2001.

Finally, we also controlled for differences in the technological regimes of the sector in which firms operate (Agarwal and Audretsch, 2001). In this perspective we classified the sampled firms according to Pavitt's taxonomy (1984) (see Appendix A for the details) in science-based, supplier-dominated, specialized suppliers and scale-intensive.

>>> INSERT Table 2 ABOUT HERE <<<

Table 2 provides a synthesis of the variables that we will use in the empirical analysis.

3.3 Methodology

In order to evaluate the effects of the structural properties of firms' knowledge bases on post-entry performances we focus on the survival time of the firm which is a duration variable. If T indicates the number of months that our firms have survived up to December 2011, then we can write the cumulative distribution function F of the duration time T as follows:

$$F(t) = P(T \leq t), \qquad t \geq 0 \tag{1}$$

This specification gives the probability that the duration T is less than or equal to t. In other words, this function represents the probability that a firm exits the population before t months after December 2001.

The survival function is then defined as:

$$S(t) = 1 - F(t) = P(T > t) \quad (2)$$

Which represents the probability that a firm survives *t* months after December 2001.

The analysis is articulated in two steps. First of all we check the extent to which differences in survival rates in sampled firms can be explained by the ability to successfully undertake knowledge creation activities. In this perspective we calculated the empirical survival function by using the life-table approach (Kalbfleish and Prentice, 1980) and then estimated the survival functions for different categories of firms on the basis of their innovative performance. In this step, we simply distinguished between innovators and non-innovators by generating a dummy which takes value 1 if the knowledge capital stock of the firm is different from 0. We also performed statistical test of equality of survival distributions across the different categories of firms, and in particular the log-rank, the Wilcoxon and the Cox test.

While this former step allows us to assess whether knowledge assets may provide an explanation of post-entry performances, the purpose of this paper goes well beyond this. Once we acknowledge the role of knowledge creation activities, we aim at investigating whether differential survival patterns within the innovators subsample are explained by the structural properties of firms knowledge base. This leads us to estimate a duration model in which the survival time is function of a vector of covariates which consists of the measures described in Sections 3.2 and 3.3.

In the literature different empirical strategies have been followed to empirically estimate the determinants of differential survival rates. A number of papers have adopted traditional estimation models for binary categorical variables. Audrestch (1991) implements a logit analysis on survival rates, while Helmers and Rogers (2010) adopt a probit regression on the probability of exit of firms. Fritsch et al. (2007) use an OLS estimation on survival rates, in a framework better suited to tobit regression. On a different perspective, some other papers have instead made explicit use of duration models. Audretsch and Mahmood (1995) and Agarwal and Audretsch (2001) implemented a Cox proportional hazard regression, which is based on hazard rations. Cefis and Marsili (2006) used a parametric approach based on accelerated time models.

In this paper we will follow this latter approach, as the test based on Schoenfeld residuals suggests that our data violates the proportional hazard assumption. The accelerated time models assume a linear form for the effects of the explanatory variables on survival time and also for the underlying survival function. The data in our dataset, as is often the case in duration models, are well suited to be represented by a lognormal distribution.

The accelerated time model estimated with survival time distributed as a lognormal is given by:

$$\ln(T) = X\beta + \sigma\varepsilon \tag{3}$$

Where T is the survival time, X is the matrix of explanatory variables (see Table 2), β is the vector of the coefficients to be estimated and ε is the vector of the residuals assumed to be normally distributed. Since all the explanatory variables are calculated in logarithms, the coefficients β of the model can be interpreted as the elasticities of the covariates on the expected survival time. The parameters are estimated by maximum likelihood.

4 Results

4.1 Descriptive Statistics

Before presenting the results of the econometric estimations, we compare the general characteristics of the sampled firms, with a special focus on the distinction between innovating and non-innovating firms.

Table 3 presents the descriptive statistics for all the explanatory variables. Of course, non-innovating firms do not display any statistics for what concerns the knowledge-related variables, and knowledge capital is null. Let us recall that size is measured in term of sales. The data show that on average the size of non-innovating firms is close to the overall mean, although significantly lower than the average size of innovating firms. A similar evidence concerns also firms' age. Indeed non-innovating firms show values very close to the overall figure. Moreover, non-innovating firms in our sample are on average younger than innovating firms.

>>> INSERT Table 3 ABOUT HERE<<<

We also compare the two group of firms for what concerns their sectoral distribution in terms of Pavitt's sectors (Table 4). The two groups are clearly different also in this respect. Indeed the bulk of non-innovating firms (59.24%) is in the suppliers dominated sector, followed by the science based sector (20.63%), while most of innovating firms (49.71%) are in the science-based sector, followed by the scale and information intensive sector (22.11%).

>>>INSERT Table 4 ABOUT HERE <<<

Table 5 reports the correlation matrix for survival time (number of months a firm survives since December 2001), the knowledge-related variables and the control variables. As expected age is positively and significantly correlated with firm's survival. The same applies also for knowledge capital. This latter also shows a high and significant correlation with the other knowledge-related variables. For this reason we use knowledge capital only in the first step of our empirical strategy, which is aimed at assessing the impact of innovative behaviour on survival. Actually in the second step we focus on the impact on knowledge-related variable on the survival on the sub-sample of innovating firms.

>>> INSERT Table 5 ABOUT HERE <<<

4.2 Univariate and graphical analysis

In order to assess the impact of the properties of knowledge base on the likelihood of firms survival we first investigate whether engaging in knowledge creating activities can explain to some extent the variety in post-entry performances. In this direction, we report in Table 6 the life-table estimates of survival rates of the two groups, i.e. innovators and non-innovators. The two groups are identified on the basis of the values of the knowledge capital stock variable. In particular innovators are those firms for which the logarithm of knowledge capital is higher than zero. The table shows that at the end of the observed period the percentage of firms that exited is about 27% for the non-innovators and about 24% for the innovators. At the end of the period the survival rate of non-innovators was about 5% lower than that of innovators. We can also notice that for most of the observed years the survival rate of non-innovators is lower than that of innovators.

>>> INSERT Table 6 ABOUT HERE <<<

An alternative way to grasp these different patterns can be represented by the plot of the Kaplan-Meier survivor function (Figure 2). The Kaplan-Meier estimator (Kaplan and Meier, 1958) is a simple frequency non-parametric estimator, and as such it does not make any ex-ante assumption about the distribution of exit times. The estimator is given by:

$$\hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right) \tag{4}$$

Where n_i denotes the number of firms in the risk set at t_i and d_i the number of exits at t_i. The product is over all observed exit times that are less than or equal to t.

>>> INSERT Figure 2 ABOUT HERE <<<

The graph shows that actually the function for innovators (dashed line) is above the one representing the survival rates of non-innovators (continuous line). Finally, we also investigated whether the differences between the survival functions are statistically significant. To this purpose we implemented three statistical tests reported in Table 7. It can be seen that in all of the tests the differences between innovators and non-innovators are significant at the 1% level.

>>> INSERT Table 7ABOUT HERE <<<

The analysis conducted so far suggests that engaging in technological activities may enhance the probabilities for firms to survive. In this respect, the results are much in line with the existing literature on the subject. The analysis of the effect of the properties of firms' knowledge bases requires however the adoption of a regression framework allowing for the evaluation of the effects of the different covariates.

4.3 Multivariate analysis

In this section we provide an empirical account of the relationship between firms' characteristics and firm survival. To this purpose we estimated a parametric survival model with lognormal distribution, including a set of firm-specific covariates. Following Cefis and Marsili (2005), we estimated different specifications of the model by adopting a three-stage hierarchical procedure.

The first specification (Model 1) is aimed at assessing the contribution of knowledge capital stock. This specification not only allows for a direct comparison between innovators and non-innovators, but also an assessment of the effects of differential endowments in terms of knowledge capital. The base model 1.A takes into account firm size and age as control variables, while in the model 1.B we also include the squared terms on size and age. Finally model 1.C also includes the control for firms' growth. The second specification is aimed at assessing the differential impact of the properties of the knowledge base. This estimation is therefore restricted to the sub-sample of innovating firms, as according to our definition non-innovators does not possess any measurable knowledge base. The procedure aimed at assigning a zero values for the properties of the knowledge base would not work indeed, as the zero value does not stand for the absence of the property. For example, knowledge coherence may take both positive and negative values, i.e. zero is in the interval of possible values and would signal the presence of an average level of integration of the knowledge base. In the same vein, a zero value for cognitive distance would imply a knowledge base built upon the recombination of very similar technologies. The hierarchy of Model 2 is the same as the one of Model 1. Model 2.A includes only size and age as firm control variables, Model 2.B extends the covariates list to the squared terms of size and age and finally Model 2.C also accounts for the firms' differential growth rates. Sectoral dummies are included in all of the specifications.

>>> INSERT Table 8 ABOUT HERE <<<

The results of the econometric estimations are reported in Table 8. As for Model 1, the coefficients for size and age are also quite robust across the three specifications. In particular, size shows a negative and significant coefficient, while age a positive and significant one. The evidence on size is not in line with the literature. However, when we include the squared effect on size, the situation gets clearer. Indeed the sign on the coefficients suggest the existence of a U-shaped relationship between size and survival rate, according to which the likelihood to survive begins to increase beyond a critical value of firms' size. This is also reflected by the fact that innovators in our sample show an average size significantly higher than non-innovators. The coefficient on the squared age term suggests instead the existence of an inverted U-shaped relationship with firm survival, as indicated in some previous analyses (Audretsch and Mahmood, 1994; Wagner, 1994). The coefficient on knowledge capital turns out to be positive and statistically significant across the three specifications. This is consistent with what we have observed in the univariate analysis and supports the idea that the more a firm commits resources to the development of their knowledge base the higher are its chances to survive. However this is only part of the story.

In the theoretical framework articulated in Section 2 we indeed formulated the hypothesis that not only knowledge matters from a quantitative viewpoint, but also from a qualitative one. By adopting a collective knowledge approach, we propose to characterize the structure of firms' knowledge base on the basis of three properties, i.e. knowledge coherence, cognitive distance and knowledge variety. The former provides of a measure of the degree of complementarity across the bits of knowledge that are combined together in the knowledge base. The second provides a measure of the extent to which the combined knowledge bits are dissimilar while the latter refers to the degree of technological differentiation.

The idea is that the contribution of technological activities to survival changes according to the kind of search strategies followed by innovating firms. The results of Model 2 provide full support to our hypotheses. The coefficient on knowledge coherence, like the one on knowledge variety, is indeed positive and significant across the three specifications. On the opposite, cognitive distance shows a negative and significant coefficient.

This is consistent with the idea that the higher the degree of technological variety, the higher the success of innovation activities and hence the higher the profitability of the firm. This affects also the likelihood of firms' survival. However, the positive effect of knowledge variety is not related with the combination of knowledge bits, no matter which technological domain they come from. On the contrary, the coefficients on knowledge coherence and cognitive distance suggest that firms searching in areas of the technological landscape with which they are more familiar, have higher chances to survive. In other words, the direction of knowledge efforts with respect to the competences cumulated over time matters in shaping the post-entry performance of firms.

5 Conclusions and Policy Implications

One basic issue in the paper, consistent with the literature, is that survival likelihood should not be related only to age and size. Age and size are firms characteristics that allow for a distinction between the formative stages of the industry and the more mature ones, explaining why and how firms maintain themselves in the market over a medium or longer time period. However, age and size are not the only dimensions to be taken into account. The idea that firms engage in product innovation in the explorative stages, and succeed to survive in the exploitation phases as they move progressively or more drastically into process innovation has been also debated as a crucial issue. In that perspective, the main outcome of the paper, contributing to the advancement of the literature, is that we get to know how this intentional process of engaging innovation matters for firms survival in the dynamics in the industry. Summing up, the value of the theoretical view developed in this paper is thus the following: using such an analysis we grasp the way in which firms influence the development of the industry, while most of the existing contributions rest on how the innovation strategies are shaped by the emergence, growth, maturity and decline of sectors.

The recombinant and collective approach chosen to develop that paper provides a general framework to understand how search processes along the technological landscape can be computed by innovative firms, taking into account that this search process may be more or less diversified, coherent, and based on the combination of more or less distant bits of knowledge. On the issue of survival, the rationale behind this theoretical framework is that firms engaging an innovation process necessarily have to combine different pieces of knowledge in the perspective or creating new competences, and that this combination needs to be coherent over time with the ability of firms to generate complementarities, and has to be marked by a rather smooth development of competences rather than by radical and dissimilar associations. Our empirical study sustains these research assumptions, as it shows that there is a positive and significant relationship between variety and survival, a positive and significant relationship between coherence and survival, and a negative and significant relationship between cognitive distance and survival. This means that if they wish to survive longer, firms have thus to maintain a large variety of technological classes in their patent portfolios, controlling for complementarity effects and avoiding ruptures in the development of the knowledge base.

The analysis carried out in this paper allows to deriving some implications for technology policy. The latter cannot be implemented in an unselective way, but it should rather take into account the differential conditions of the technological environment in which firms operate. Technology policy should therefore be customized according to the relative stage of development of the technological trajectory. On the one hand, firms involved in the exploration of untried areas of the technology landscape are likely to face a more turbulent and uncertain environment and therefore more exposed to failure. These firms seem to need a specific support at the policy level. As a consequence, the collective dimension of knowledge generation calls for the full appreciation of the systemic character of technological innovation and therefore demands the adoption of a systemic perspective in technology policy. In this direction, technology policies should be oriented towards the promotion of clusters or technological platforms enabling firms in uncertain technological environments to access external technology competences and enhance exploration activities. A particular attention ought to be paid in this respect not only to key dimensions like geographical proximity or firms' characteristics, but also to technological proximity amongst interacting agents. On the other hand, demand-driven innovation policies aiming at fostering the development of local competitive advantages should be targeted towards the support of innovating firms operating in stable and established technological environments and, thus, involved in the exploitation of their accumulated competences. In these latter the probability to exploit successful innovations is indeed much higher, and so is the likelihood to achieve technology based economic growth.

6 References

[42] Acs, Z.J., Anselin, L. and Varga, A., 2002, "Patents and innovation counts as measures of regional production of new knowledge", *Research Policy*, 31, 1069-1085.

[58] Agarwal, R., 1991. "Survival of firms of the product life cycle", *Southern Economic Journal*, 63, 571-584.

[21] Agarwal, R. and Audretsch, D., 2001. "Does Entry Size Matter? The Impact of the Life Cycle and Technology on Firm Survival", *Journal of Industrial Economics*, 49(1), 21-43.

[2] Audretsch, D., 1991, "New-Firm Survival and the Technological Regime", *The Review of Economics and Statistics*, 73(3), 441-50.

[11] Audretsch, D., 1995. "Innovation, growth and survival", *International Journal of Industrial Organization*, 13(4), 441-457.

[3] Audretsch, D. and Mahmood, T. 1994, "Firm Selection and Industry Evolution: The Post-entry Performance of New Firms", *Journal of Evolutionary Economics*, 4(3), 243-60.

[62] Audretsch, D. and Mahmood, T., 1995, "New Firm Survival: New Results Using a Hazard Function", *The Review of Economics and Statistics*, 77(1), 97-103.

[12] Baldwin, J.R., 1995, *The dynamics of industrial competition: A North American perspective*, Cambridge University Press, Cambridge.

[27] Banbury, C. M. and W. Mitchell, 1995, "The effects of introducing important incremental innovation on market share and business survival' , *Strategic Management Journal*, 16, 161-182.

[54] Breschi, S., Lissoni, F., and Malerba, F., 2003, "Knowledge relatedness in firm technological diversification". *Research Policy*, 32, 69-97.

[57] Bruderl and Schussler, 1990, "Organizational Mortality: The Liabilities of Newness and Adolescence", *Administrative Science Quarterly*, 35, 530–547.

[9] Caves, Richard E., 1998, "Industrial Organization and New Findings on the Turnover and Mobility of Firms", *Journal of Economic Literature*, 36(4), 1947-1982.

[24] Caves, R. and Porter, M., 1977, "From Entry Barriers to Mobility Barriers," *Quarterly Journal of Economics*, 91, pp. 241-261.

[4] Cefis, E., and Marsili, O., 2005, "A Matter of Life and Death: Innovation and Firm Survival", *Industrial and Corporate Change*, 14(6), 1167-1192.

[5] Cefis, E., and Marsili, O., 2006, "Survivor: the Role of Innovation in Firms' Survival", *Research Policy*, 35(5), 626-641.

[26] Christensen, C., Suarez, F., and Utterback, J., 1998, "Strategies for survival in fast-changing industries", *Management Science*, 44, 207-S220.

[37] Dosi G., 1982, "Technological paradigms and technological trajectories", *Research Policy*, 11, 147–162.

[10] Dunne, T., Roberts, M. and Samuelson, L., 1988. "Patterns of Firm Entry and Exit in U.S. Manufacturing Industries", RAND Journal of Economics, 19(4), 495-515.

[19] Ericson, R. and Pakes A., 1995, "Markov-Perfect Industry Dynamics: A Framework for Empirical Work", *Review of Economic Studies*, 62, 53-82.

[55] Engelsman, E.C. and van Raan, A.F.J, 1994, A patent-based cartography of technology, *Research Policy*, 23, 1-26.

[17] Evans, D., 1987. "Tests of Alternative Theories of Firm Growth", *Journal of Political Economy*, 95(4), 657-74.

[36] Fleming, L., 2001, "Recombinant Uncertainty in Technological Search", *Management Science*, 47(1), 117-132.

[56] Fleming, L., Mingo, S. and Chen, D., 2007, "Collaborative brokerage, generative creativity and creative success", *Administrative Science Quarterly*, 52, 443-475.

[45] Frenken, K., and Nuvolari, A., 2004. “The early development of the steam engine: an evolutionary interpretation using complexity theory”, *Industrial and Corporate Change*, 13, 419–450.

[61] Fritsch, M., Brixy, U. and Falk, O., 2006, “The Effect of Industry, Region, and Time on New Business Survival – A Multi-Dimensional Analysis”, *Review of Industrial Organization*, 28, 285-306.

[15] Geroski, P., 1995, “What Do We Know about Entry?”, *International Journal of Industrial Organization*, 13, 421-440.

[29] Griliches, Z., 1979, “Issues in Assessing the Contribution of Research and Development to Productivity Growth ”, *Bell Journal of Economics*, 10(1), 92-116.

[41] Griliches, Z., 1990, “Patent statistics as economic indicators: a survey”, *Journal of Economic Literature*, 28, 1661-1707.

[18] Hall, B. H., 1987, “The relationship between firm size and firm growth in the US manufacturing sector”, *The Journal of Industrial Economics*, 35(4), 583-606.

[43] Hall, B.H., Griliches, Z., Hausman, J.A., 1986, “Patents and R&D: is there a lag?”, *International Economic Review*, 27, 265–283.

[6] Helmers C., and Rogers, M., 2010, “Innovation and the Survival of New Firms in the UK ”," *Review of Industrial Organization*, 36(3), 227-248.

[53] Jaffe, A.B, 1989, “Characterising the technological position of firms, with application to quantifying technological opportunity and research spillovers”, *Research Policy*, 18, 87–97.

[52] Jaffe, A.B., 1986, “Technological opportunity and spillovers of R&D: evidence from firms’ patents, profits, and market value”, *American Economic Review*, 76, 984–1001.

[16] Jovanovic, B., 1982, “Selection and evolution of industry”, *Econometrica* 50 (3), 649–670.

[60] Kalbfleisch JD, Prentice RL., 1980, *Statistical Analysis of Failure Time Data*, New York: Wiley.

[62] Kaplan, E. L. and Meier P., 1958, “Non-parametric Estimation from Incomplete Observations”. J. Am. Stat. Assoc., 53, 457–481.

[32] Kaufman, S., 1993. *Origins of Order: Self-organization and Selection in Evolution*, Oxford University Press, Oxford.

[22] Klepper, S., 1996. “Entry, Exit, Growth, and Innovation over the Product Life Cycle”, *American Economic Review*, 86(3), 562-83.

[23] Klepper, S., 1997, “Industry Life Cycles”, *Industrial and Corporate Change*, 6(1), 145-81.

[7] Krafft, J., Quatraro, F., Saviotti, P.P., 2009. Evolution of the knowledge base in knowledge intensive industries, LEI and BRICK Working Paper no. 06/2009.

[30] Mansfield, E., 1980, “Basic research and productivity increase in manufacturing”, *American Economic Review*, 70, 863-73.

[28] March, J., 1991, “Exploration and exploitation in organizational learning”, *Organization Science*, 2(1), 71-87.

[14] Mata, J., Portugal, P., 1994, “Life duration of new firms”, *Journal of Industrial Economics*, 42 (3), 227–246.

[20] Nelson, R.R., Winter, S.G., 1982. *An Evolutionary Theory of Economic Change*. Belknap Press of Harvard University Press, Cambridge, MA.

[48] Nesta, L., 2008, “Knowledge and productivity in the world’s largest manufacturing corporations”, *Journal of Economic Behavior and Organization*, 67, 886–902.

[46] Nesta, L., Saviotti, P., 2005, “Coherence of the knowledge base and the firm’s innovative performance: evidence from the U.S. pharmaceutical industry”, *Journal of Industrial Economics*, 53, 123–142.

[47] Nesta, L., Saviotti, P.P., 2006, “Firm knowledge and market value in biotechnology”, *Industrial and Corporate Change*, 15, 625–652.

[35] Nightingale, P., 1998, “A cognitive model of innovation”, *Research Policy*, 27(7), 689-710.

[51] Nooteboom, B., 2009, *A cognitive theory of the firm; Learning, governance and dynamic capabilities*, Cheltenham UK: Edward Elgar.

[59] Pavitt, K., 1984, “Sectoral patterns of technical change: towards a taxonomy and a theory”, *Research Policy*, 13: 343–373.

[40] Pavitt, K., 1985, “Patent statistics as indicators of innovative activities: possibilities and problems”, *Scientometrics*, 7, 77–99.

[25] Perez, S., Llopis, A. and Llopis, J., 2004, The determinants of survival of Spanish manufacturing firms, *Review of Industrial Organization*, 25, 251-273.

[8] Quatraro, F, 2012, *The Economics of Structural Change in Knowledge,* London and New York, Routledge.

[1] Santarelli, E. and Vivarelli, M., 2007. “Entrepreneurship and the process of firms’ entry, survival and growth”, *Industrial and Corporate Change*, 16(3), 455-488.

[44] Saviotti, P.P., 1988, “Information, variety and entropy in technoeconomic development”, *Research Policy*, 17, 89–103.

[33] Saviotti, P.P., 2004, “Considerations about the production and utilization of knowledge”, *Journal of Institutional and Theoretical Economics*, 160, 100-121.

[34] Saviotti, P.P., 2007, “On the dynamics of generation and utilisation of knowledge: The local character of knowledge”, *Structural Change and Economic Dynamics*, 18, 387-408.

[39] Soete, L. and Patel, P., 1985, “Recherche-developpement, importations de technologie et croissance économique: une tentative de comparaison internationale”, *Revue Economique*, 36, 975-1000.

[50] Stirling, A., 2007, “A general framework for analysing diversity in science, technology and society”, *Journal of the Royal Society Interface*, 4 (15). pp. 707-719.

[13] Sutton, J., 1997, “Gibrat's legacy”, *Journal of Economic Literature*, 35(1), 40-59.

[49] Teece, D., Rumelt, R., Dosi, G. And Winter, S., 1994, “Understanding Corporate Coherence: Theory and Evidence”, *Journal of Economic Behavior and Organization*, 23, 1–30.

[38] Thoma, G., Torrisi, S., Gambardella, A., Guellec, D., Hall, B.H. and Haroff, D., 2010, Harmonizing and combining large datasets – an application to firm-level patent and accounting data, NBER Working Paper 15851.

[64] Wagner, J., 1994, “The post-entry performance of new small firms in German manufacturing industries”, *The Journal of Industrial Economics*, 42(2), 141-154.

[31] Weitzman, M. L., 1998, “Recombinant growth”, *Quarterly Journal of Economics*, 113, 331-360.

APPENDIX A

Table A1 – Correspondence between Pavitt's groups and NACE rev. 2 classification

SCIENCE BASED	**Nace rev2**
Chemicals	20, 21
Office machinery	28
Manufacture of radio, television and communication equipment and apparatus	26
Manufacture of medical, precision and optical instruments, watches and clocks	26
Communications	58, 59, 63
Computer and related activities	62
Research and development	72
SCALE AND INFORMATION INTENSIVE	
Pulp, paper & paper products	17
Printing & publishing	18
Mineral oil refining, coke & nuclear fuel	19, 81
Rubber & plastics	22, 89
Nonmetallic mineral products	23
Basic metals	24
Motor vehicles	29
Financial intermediation, except insurance and pension funding	64
Insurance and pension funding, except compulsory social security	65
Activities auxiliary to financial intermediation	66
SPECIALISED SUPPLIERS	
Mechanical engineering	71
Manufacture of electrical machinery and apparatus n.e.c.	27
Manufacture of other transport equipment	30
Real estate activities	68, 41
Renting of machinery and equipment	77
Other business activities	82
SUPPLIERS DOMINATED	
Food, drink & tobacco	10, 11, 12
Textiles	13
Clothing	14
Leather and footwear	15
Wood & products of wood and cork	16
Fabricated metal products	25
Furniture, miscellaneous manufacturing; recycling	31, 32, 33
Sale, maintenance and repair of motor vehicles and motorcycles; retail sale of automotive fuel	42, 43, 45
Wholesale trade and commission trade, except of motor vehicles and motorcycles	46
Retail trade, except of motor vehicles and motorcycles; repair of personal and household goods	47
Hotels & catering	55, 56
Inland transport	49
Water transport	50
Air transport	51
Supporting and auxiliary transport activities; activities of travel agencies	52

Figure 1 - Distribution of patent applications across firms

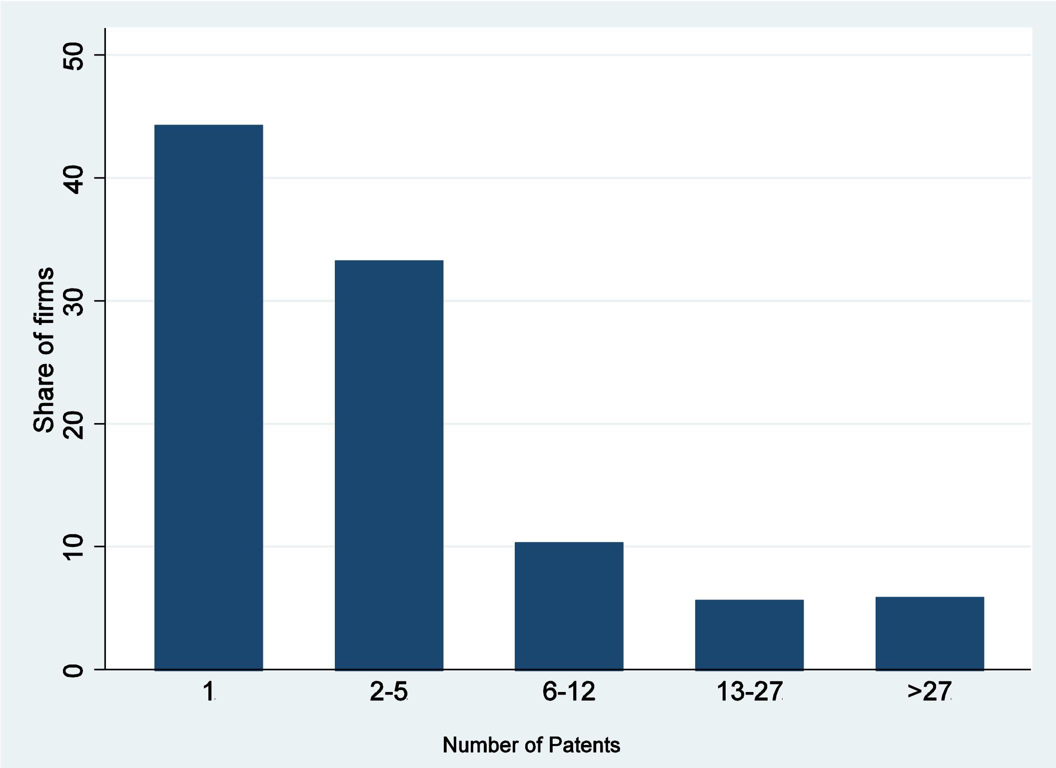

Figure 2 - Comparison of survival function between innovators and non-innovators

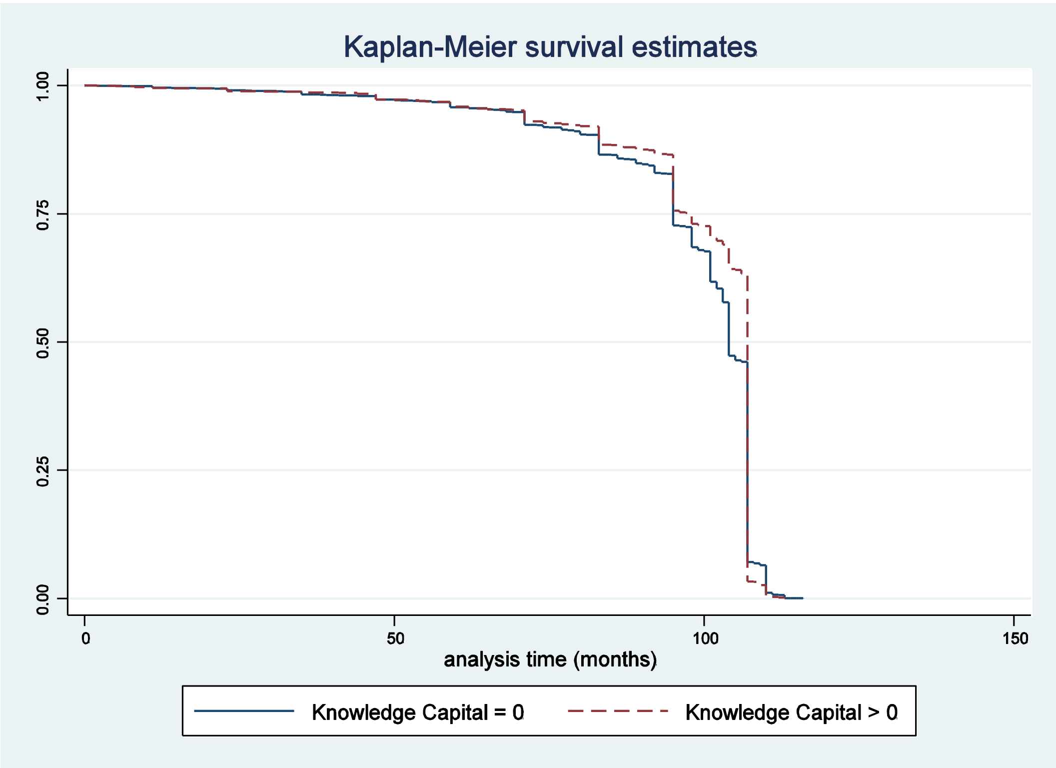

Table 1 – Sectoral distribution of firms in the dataset as compared to the distribution at the national level.

		Sample			France		
Nace Code (rev 2)	Industry	Freq.	%	Cum	Freq.	%	Cum
5	Mining of coal and lignite	5	0,006	0,006	10	0,001	0,001
6	Extraction of crude petroleum and natural gas	26	0,033	0,039	61	0,009	0,010
7	Mining of metal ores	12	0,015	0,054	56	0,008	0,019
8	Othermining and quarrying	844	1,069	1,124	1885	0,277	0,296
9	Mining support service activities	16	0,020	1,144	49	0,007	0,303
10	Manufacture of foodproducts	5998	7,598	8,741	59488	8,742	9,045
11	Manufacture of beverages	1183	1,498	10,240	2725	0,400	9,446
12	Manufacture of tobaccoproducts	4	0,005	10,245	5	0,001	9,447
13	Manufacture of textiles	1132	1,434	11,679	4180	0,614	10,061
14	Manufacture of wearingapparel	1153	1,460	13,139	8055	1,184	11,245
15	Manufacture of leather and related products	382	0,484	13,623	1946	0,286	11,531
16	Manufacture of wood and of products of wood and cork	21	0,027	13,650	9139	1,343	12,874
17	Manufacture of paper and paper products	851	1,078	14,728	1477	0,217	13,091
18	Printing and reproduction of recorded media	2392	3,030	17,758	16234	2,386	15,476
19	Manufacture of coke and refined petroleum products	61	0,077	17,835	88	0,013	15,489
20	Manufacture of chemicals and chemical products	1369	1,734	19,569	2874	0,422	15,912
21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	273	0,346	19,915	475	0,070	15,982
22	Manufacture of rubber and plastic products	2429	3,077	22,992	4747	0,698	16,679
23	Manufacture of other non-metallic mineral products	1789	2,266	25,258	8955	1,316	17,995
24	Manufacture of basic metals	571	0,723	25,981	1095	0,161	18,156
25	Manufacture of fabricated metal products, except machinery and equipment	8053	10,201	36,182	19477	2,862	21,018
26	Manufacture of computer, electronic and optical products	1325	1,678	37,860	4063	0,597	21,616
27	Manufacture of electricalequipment	108	0,137	37,997	2656	0,390	22,006
28	Manufacture of machinery and equipment n.e.c.	309	0,391	38,388	7928	1,165	23,171
29	Manufacture of motor vehicles, trailers and semi-trailers	931	1,179	39,568	2146	0,315	23,486
30	Manufacture of other transport equipment	326	0,413	39,980	995	0,146	23,633
31	Manufacture of furniture	1195	1,514	41,494	14200	2,087	25,719
32	Othermanufacturing	1413	1,790	43,284	16136	2,371	28,091
33	Repair and installation of machinery and equipment	2727	3,454	46,738	22565	3,316	31,407
41	Construction of buildings	5126	6,493	53,231	44007	6,467	37,874
42	Civil engineering	2032	2,574	55,805	6449	0,948	38,822
43	Specialised construction activities	32321	40,941	96,746	387333	56,923	95,745
58	Publishingactivities	2455	3,110	99,856	13506	1,985	97,730
59	Motion picture, video and television programme production	114	0,144	100,000	15448	2,270	100,000
		78946	100		680453	100	

Note: National-level data have been drawn from Eurostat Structural Business Statistics and refer to the year 2008.

Table 2–Variable used in the empirical analysis

Variable	Measure	Time
Survival time	Elapsed months since Jan 2002 to exit	
Age	Logarithm of firms age since its foundation	Evaluated at 2002
Age^2	Square of Age	Evaluated at 2002
Size	Logarithm of firm sales	Evaluated at 2002
$Size^2$	Square of Size	Evaluated at 2002
Size x Age	Product between Age and Size	Evaluated at 2002
Kn. Stock	Knowledge Capital Stock (PIM)	Average value on the period 1997-2002
Kn. Coherence	Coherence of the knowledge base	Average value on the period 1997-2002
Cognitive Distance	Inverse of technological proximity	Average value on the period 1997-2002
Knowledge Variety	Multidimensional Information Entropy	Average value on the period 1997-2002

Table 3 - Descriptive Statistics

	N	*mean*	*sd*	*kurtosis*	*skewness*	*median*
Non innovators						
Kn. variety						
Cognitive Distance						
Kn. Coherence						
Kn. Capital	73333	0	0	.	.	0
Age	70664	5.020	1.064	2.985	-0.285	5.050
Size	63909	7.378	1.281	5.523	0.620	7.196
Growth	38720	0.059	0.384	67.000	3.358	0.028
Innovators						
Kn. variety	536	0.644	0.891	2.419	-0.329	0.717
Cognitive Distance	844	-0.553	0.305	42.709	-5.478	-0.497
Kn. Coherence	832	2.377	0.514	3.958	1.152	2.242
Kn. Capital	1529	2.137	1.344	4.330	1.097	1.828
Age	1511	5.454	0.962	3.153	-0.506	5.529
Size	1420	10.064	1.953	3.747	-0.029	10.064
Growth	1068	0.012	0.429	61.226	-1.125	0.010
Total						
Kn. variety	539	0.639	0.892	2.400	-0.317	0.711
Cognitive Distance	891	-0.557	0.314	38.384	-5.210	-0.499
Kn. Coherence	832	2.377	0.514	3.958	1.152	2.242
Kn. Capital	74862	0.044	0.358	128.003	10.304	0.000
Age	72175	5.029	1.063	2.983	-0.290	5.050
Size	65329	7.437	1.357	5.865	0.815	7.224
Growth	39788	0.058	0.386	66.879	3.191	0.027

Note: all variables are in logarithm

Table 4 - Sectoral Distribution of Sampled Firms

	Non-Innovators			Innovators		
	Freq.	Percent	Cum.	Freq.	Percent	Cum.
Scale and information intensive	8,646	11.79	11.79	338	22.11	22.11
Science based	15,126	20.63	32.42	760	49.71	71.81
Specialised suppliers	6,117	8.34	40.76	142	9.29	81.1
Suppliers dominated	43,444	59.24	100	289	18.9	100
Total	73,333	100		1,529	100	

Table 5 - Correlation Matrix

	Survival time	Kn. Variety	Cognitive Distance	Kn. Coherence	Kn. Capital	Age	Size	Growth
Survival time	1							
Kn. Variety	**0.129**	1						
	(0.003)							
Cognitive Distance	-0.012	0.062	1					
	(0.726)	(0.151)						
Kn. Coherence	0.007	**-0.113**	**-0.210**	1				
	(0.837)	(0.009)	(0.000)					
Kn. Capital	**0.020**	**0.722**	**0.081**	-0.034	1			
	(0.000)	(0.000)	(0.016)	(0.322)				
Age	**0.033**	0.062	0.053	0.005	**0.063**	1		
	(0.000)	(0.153)	(0.116)	(0.887)	(0.000)			
Size	-0.003	**0.466**	**0.113**	-0.019	**0.302**	**0.334**	1	
	(0.401)	(0.000)	(0.001)	(0.590)	(0.000)	(0.000)		
Growth	-0.009	-0.046	**-0.084**	**0.092**	**-0.020**	**-0.187**	**0.021**	1
	(0.071)	(0.358)	(0.035)	(0.024)	(0.000)	(0.000)	(0.000)	

Note: p-values in parentheses

Table 6 - Survival rates by sample

Year	Non-innovators	Innovators
0	1	1
1	99.58	99.54
2	99.08	98.89
3	98.28	98.63
4	97.26	97.25
5	95.81	95.88
6	92.34	93
7	86.54	88.49
8	72.71	75.6
Number of firms	73333	1529
Percentage of failure	27.29	24.4

Note: Life-table estimates of survival rates

Table 7 - Test of equality of survival functions (innovators vs non-innovators)

Test	Chi-square	Pr > Chi-square
Log-rank	29.94	0.00
Wilcoxon	35.90	0.00
Cox	16.62	0.00

Table 8 – Results of the Econometric Estimations

	Model 1			Model 2		
Variables	A	B	C	A	B	C
Kn. Capital	0.0146***	0.0118***	0.0125***			
	(0.00315)	(0.00316)	(0.00267)			
Kn. Coherence				0.0371*	0.0379**	0.0314*
				(0.0210)	(0.0190)	(0.0175)
Kn. Variety				0.0170**	0.0162**	0.0146
				(0.00813)	(0.00799)	(0.00915)
Cognitive Distance				-0.0799**	-0.0856**	-0.0807*
				(0.0323)	(0.0352)	(0.0444)
Age	0.00785***	0.0327***	0.0212**	0.00238	-0.131	-0.0602
	(0.00112)	(0.00834)	(0.0103)	(0.00988)	(0.105)	(0.0953)
Age^2		-0.00281***	-0.00214*		-0.00176	0.00139
		(0.000826)	(0.00124)		(0.00632)	(0.00612)
Size	-0.00514***	-0.0216***	-0.0267***	0.00732	-0.0478	-0.0325
	(0.00103)	(0.00557)	(0.00842)	(0.00472)	(0.0309)	(0.0226)
Size^2		0.000941***	0.00100**		-0.000800	0.000372
		(0.000277)	(0.000405)		(0.00196)	(0.00107)
Sales x Age		0.000296	0.00117		0.0134	0.00523
		(0.000669)	(0.00108)		(0.0107)	(0.00431)
Growth			-0.00120			-0.0298
			(0.00421)			(0.0200)
Constant	4.580***	4.586***	4.633***	4.326***	5.001***	4.746***
	(0.00715)	(0.0318)	(0.0399)	(0.100)	(0.375)	(0.338)
Log-likelihood	-1.369	-1.369	-1.421	-1.710	-1.717	-1.921
Observations	64,374	64,374	39,785	503	503	404

*** p<0.01, ** p<0.05, * p<0.1